

LINGUISTIC DIVERSITY IN ONLINE SPACE - A STUDY ON SOUTH ASIAN LANGUAGES

DR. K. S. ARUL SELVAN

The Indira Gandhi National Open University, New Delhi, India

ABSTRACT

Rich linguistic diversity is hallmark of South Asian regions that comprises eight nations. Some of the languages are spoken in more than one country, and scripts of few languages are common. This shared commonality increases the cultural diversity within this region. Equally South Asian region is witnessing massive growth of Internet users barring few nations. With the affordable Internet connectivity through multiple telecommunication players, many Internet users are active in social media. These scenarios are highly congenial for any language to adapt into online platform easily. With this assumption, this article is attempting to explore the linguistic diversity and web presence of South Asian languages in online space. To measure the web presence of these languages, this article relies on three sets of data - Wikipedia articles, Google search techniques, and online algorithm. With these three indicators, an attempt would be made to measure the nature and present scenario of adaptability nature of South Asian languages in online space.

KEYWORDS: South Asian Languages, Multilingualism, Linguistic Diversity & Online Space

Received: Apr 19, 2019; **Accepted:** May 10, 2019; **Published:** Jun 08, 2019; **Paper Id.:** IJCMSJUN201919

INTRODUCTION

Since the post liberalization a sustained growth was witnessed across the South Asia, primarily in the television sector and significantly also in online media and newspapers market, particularly in the language media section. The South Asian region is popularly known for its high density of population coupled with the higher range of illiteracy and poverty, ethnic violence, terrorist activities, and restoration of democracy.

Among the eight nations - Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan and Sri Lanka - India is in dominant position due to its sheer size of market and country. Afghanistan is crippled with the internal terrorists activities which mars the growth of divergent media plurality, Bangladesh is comparatively providing better platform for media, monarchy set-up dominates in Bhutan, lack of economic viability grips the media growth in Maldives, restoration of democracy is moot point in Nepal which propelled community level media growth, Pakistan has witnessed strong growth of regional television channels, ethnic war deprived the growth of media in Sri Lanka. Lets see an overview of media scenario in these individual countries:

Afghanistan: According to BBC Media Action, there are 64 private television channels and 22 state owned channels. There are 175 radio stations. Mostly these channels are in regional languages.

Bangladesh: Once the stronghold of radio presence, now the country mostly glued to more television. According to governmental sources, there are 292 dailies, 125 weeklies and 30 monthlies published across the country. Bangla is the popular language in all these publications. There are 23 satellite channels besides large numbers of foreign channels are available in Bangladesh; particularly Indian channels are popular here.

Bhutan: State funded BBS is the only television channel, but cable TV thrives here. There are 11 newspapers (seven in English), five radio stations, and scores of magazine are available.

India: Due to the population dividend, India has got majority of young adults, which, coupled with the rising literacy rate and economic affordability led to rise of regional language dailies has got a significant reach in India. According to Indian Readership Survey 2017, eight top most circulated newspapers were from regional languages and two from English. As per Ministry of Information and Broadcasting data (December 2018) there were 906 television channels and out of which 445 were from news segments, in this segment majority of the news channels were from regional languages and hardly few English channels. According to Film Federation of India, in 2017 alone there were 1986 films certified for screening, among them Hindi, Tamil and Telugu were top film producing languages.

Maldives: Privately owned newspapers, radio and TV are thriving in Maldives, with most support comes from business magnets. However, the strong media regulatory mechanism casting shadow over the growth of media diversity in this island nation. On the other hand, there is a strong expansion of online-based news outlets as well as social media based forums.

Nepal: According to an UNESCO report, in 2013, Nepal had 3408 registered newspapers, out of which 360 were dailies, 515 radio stations and 58 television channels.

Pakistan: There are about 90 television channels, mostly private, and 130 radio stations across Pakistan. The country enjoys liberalized media licensing policy since late 1990s. Pakistan film industry, called Lollywood, produces nearly 100 films in a year. With 4800 print media outlets, out of which 430 are daily newspapers. Most of these publications are in regional languages, English restricted to urban phenomenon.

Sri Lanka: With 25 television channels and 54 radio stations, Sri Lanka has impressive media scenario, with controlled media freedom situation. In print media, newspapers are having 2.25 lakhs circulations and magazine section commands 1-lakh circulation figures. In both the platforms, Sinhala language is a dominant player. On an average 20 films are produced in a year, since 2010 there are 118 films released.

Across the South Asian countries, English is considered an elite language, media based in this language always oriented towards upper class of urban segment. Conversely regional languages are ruling the roost, mainly due to new emergent middle class segment which enjoys higher literacy and better economic affordability, they are not necessarily from urban segments but hails mostly from semi-urban areas.

THE INTERNET USERS IN SOUTH ASIA

Impressively, half of world's Internet users (51.7%) are in Asian continent. Primarily, China and India are leading counties that have highest number of users. However, in terms of prevalence of Internet among South Asian countries, there is a contrasting scenario exists. Due to liberalized economic policies of these countries, telecom sectors were allowed to operated by the private entities along with the government owned players, which has got business model coincided with the emerging semi-urban population segments that resulted more people were hooked into online. However, the large majority of rural segments in many South Asian countries are still yet to be reached into digital world.

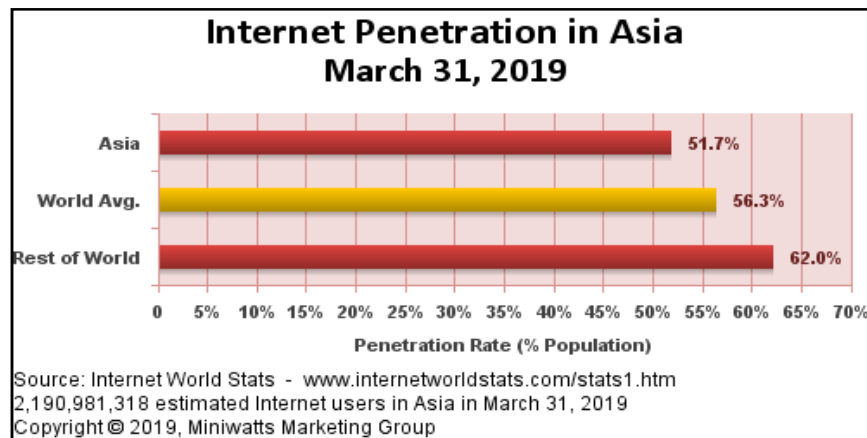


Figure 1

According to Telecom Regulatory Authority of India, in its annual report of 2017-18 it stated that India has got 493 million Internet users (as on 2019, it is 560 million - 40.9%) and 1183 million mobile subscribers. The overall tele-density was 92.84%, in that rural tele-density had risen from 56.51% to 59.01% and however urban tele-density has got reduced marginally from 171.80% to 165.90%. The tele-density figure supports argument of semi-urban population segment is significantly increasing their presence in mobile and telecom sectors. More or less this scenario exists in other South Asian countries.

Table 1: Internet Statistics as on May 2019

Country	Internet Users				Facebook Users in % as on 2019
	In 2019*		In 2015**		
Afghanistan	0.65 Million	17.6%	NA	NA	8.9%
Bangladesh	90.2 Million	54.8%	53 Million	31.9%	17.9%
Bhutan	0.39 Million	54.8%	0.25 Million	34.4%	49.8%
India	560 Million	40.9%	375 Million	30%	18.5%
Maldives	0.34 Million	75.3%	0.23 Million	58.5%	71.7%
Nepal	16 Million	54.1%	5.7 Million	18.1%	30.4%
Pakistan	44.6 Million	21.8%	29 Million	14.6%	15.3%
Sri Lanka	7.2 Million	34.1%	5.6 Million	25.8%	25.4%

** Based on April 2015 data

Internet penetration has varied presence across the region; Afghanistan (17.6%) and Pakistan (21.8%) are having the least level of Internet presence. However, both these countries have increased their users' size significantly in the last five years time (Table 1). Maldives is having highest level of presence consistently (75.3%). Bangladesh and Bhutan having equal share of Internet users (54.8%), Nepal is marginally lower than Bhutan with the 54.1% and modestly better condition in Sri Lanka (34.1%).

As per Economic Times news report (October, 2018), "according to a recent Google-KPMG report, India has 234 million Indian language users online while only 175 million are English language users. The Indian language user base is poised to account for 75% of India's Internet user base by 2021. 90% of new Internet users over the next five years are projected to be Indian language users. Adding to that, the government is planning to propagate digital literacy among 60,000 rural households by 2021." The assessment of language scenario of online space with reference to Indian context, more or less applicable to other countries in the South Asian region. However, not all of them sharing equal economic equivalence with the Indian, other than that there are other disturbing socio-political and economic issues that prevents the

proliferation of internet access in this region. Particularly, Afghanistan and Pakistan are ravaged with the terrorist activities, Nepal in amidst of pro-democracy transition situation, Bhutan and Maldives are being the small country with its unique natural conditions doesn't have prospective growth of Internet prevalence. Sri Lanka and Bangladesh are currently maintaining normalcy. However, while comparing 2015 and 2019 data of Internet users, Bangladesh, Nepal, Maldives and Bhutan have increased their users base tremendously. Srilanka, Pakistan and Afghanistan has witnessed a modest growth.

According to a news report, there are 7099 living languages globally, out of which 650 from South Asian region (The Hindu, 2018). It signifies that 10% of world's living languages exist in South Asian countries. It shows the strong and diverse language presence in this region. Therefore mentioned Internet penetration statistical details of South Asian region, indicates that there is a strong and vibrant semi-urban population segment, which is more easily accessible to regional languages. Diversity of South Asian languages and emerging semi-urban population segments, which is more hooked onto online medium, are congenial atmosphere for high prosperity of these languages are adapting into digital platform. With this assumption, this article is attempting to measure the linguistic diversity of South Asian regional in online space.

ABOUT SOUTH ASIAN LANGUAGES

There are four major language families prevalent in the South Asian context - Indo-European, Dravidian, Austro-Asiatic and Tibeto-Burman. There are other language families too. Out of the popular four, Indo-European dominates the South Asian region, 80% of the population of this region speaks languages under this category. Second most popular language family of this region is Dravidian, 18%. Linguistically there are many shared characteristics among the South Asian countries - Tamil, Bengali, Nepali, Urdu are some of the languages which are common in more than one country. These shared characteristics tend to promote closer exchanges among the people cutting across the border.

The Eighth Schedule of the Indian Constitution lists twenty-two languages that are official languages. As per Census 2011 data, there are 122 languages spoken in India and 234 mother tongues exist. In Sri Lanka, Sinhala is the national language; English and Tamil are other official languages. Tamil is the official language of Tamil Nadu state of India. Dhivehi is the official language of Maldives, however English and few Indian languages are also spoken here. Bangla language is the majority one in Bangladesh, the same language is the official language of India's West Bengal State, which share border with the Bangladesh. In Pakistan, Urdu is the official language, however Punjabi is the popular one - nearly 60% of Pakistan population speaks this language. Punjabi is the official language of Punjab State in India which shares border with the Pakistan.

The traditional media is restrictive within the geographical boundaries, except the online medium, which transcends country limits. Linguistic and shared medium commonality are the focal theme of this paper which attempts to explore the language diversity of South Asian region in the online space using a well-known parameter - web presence. Under the web presence, three indicators will be utilized to measure and compare the prevalence of South Asian languages in the online space. These indicators are - Wikipedia article strength, Google Keyword search technique and a web algorithm.

MEASURING LINGUISTIC DIVERSITY

The present author had carried a similar study within the context of Indian languages. In that article author quoted Mikami et al. 2005 to assess the usage level of every language in cyberspace a] user profile b] user activity and c] web presence. Analyzing of these 3 strategies Gerrand 2007 recommends web presence because the most sensible indicator for

estimating actual language use in cyberspace (Selvan, 2015). By following established ways that to measure linguistic diversity and establish the position of chosen languages within the online platform this study used web presence methodology to measure South Asian languages within the online world.

For this article, three indicators are considered to measure the language position online:

- Wikipedia Editions in South Asian languages
- Google search techniques
- Third party on-line language measuring formula

WIKIPEDIA EDITIONS IN SOUTH ASIAN LANGUAGES

Among three datasets, the first set of data was retrieved from Wikipedia, which contains hundreds and thousands of articles on every matter. More importantly all these articles are written by the users and subsequently these articles are under review and editing. Addition and deletion within the articles are a continuous process. The details regarding the size of Wikipedia language editions, and also the articles in every South Asian language editions are accessible at the wiki statistics. All information regarding the South Asian language based Wikipedia editions were retrieved, supported this data a ranking list has been created. For purpose of this study, South Asian languages are considered that are recognized as official languages as well as prevailing in more than one country.

The retrieved data and its ranking indicates that Urdu, Hindi and Tamil has got more than one lakh number of articles in its respective wiki editions compared to other South Asian languages. While comparing 2015 and 2019 data, Tamil language has increased its article size considerably - it was around 80 thousand four years back and but in 2019 it crossed one lakh article count. Similarly Bengali language has seen rapid growth in this period; currently it has got 63 thousand articles. The other popular language based wiki editions such as Punjabi, Nepali and Sinhala has got 10,000 to 40,000 articles. Among these languages, Punjabi has witnessed a marginal decrease in number of articles from 40 thousand to 31 thousands

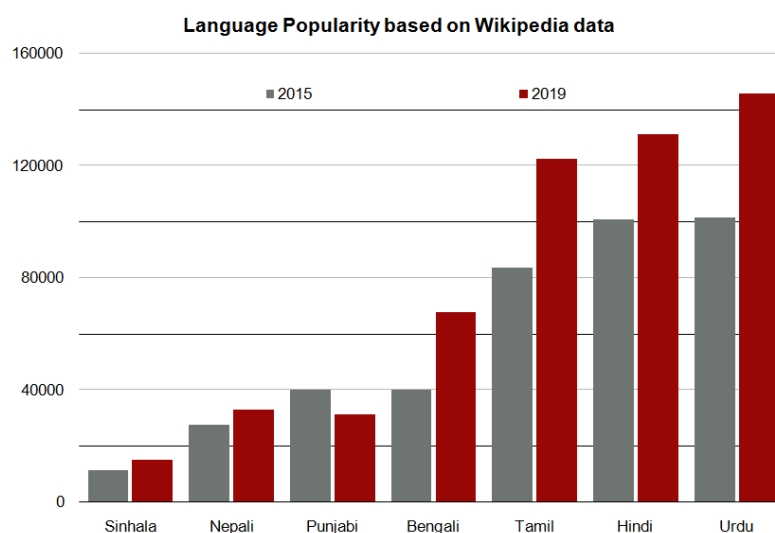


Chart 1: Source: Data Collected from the Wikipedia website

Urdu, Hindi and Tamil languages are dominating in this list of South Asian language based Wikipedia. Among three these languages, Urdu and Hindi are spoken in India and Pakistan and maintains the status of official languages in both the countries. Tamil is official language in India and Srilanka (however Tamil is an official in other countries too, but that is beyond the scope of this study). According to Ethnologue website (Table 2 - as per 2016 data) on size of world languages, Hindi has 260 million native speakers whereas Urdu has 68.6 million speakers. Almost four times lesser than Hindi language size, still Urdu made enormous level of output in Wikipedia articles. Literacy rate wise, there is a considerable differences between India and Pakistan - former's literacy rate is 62.8% and latter's rate is 54.9%. Similarly, in terms of Internet infrastructure there is a vast difference - India has 886 million mobile users and Pakistan has 127 million users. As far as Tamil is concerned, it has 67.8 million native speakers. Another interesting comparison is between Punjabi and Bangla - Punjabi has got 100 million speakers and Bangla has got 189 million. In Bangladesh, the literacy rate is 57.7% and has got 116.5 million users. Even though Sri Lanka got more than 90% literacy rate and 21 million mobile users, still Sinhala has got less than 11 thousand articles only.

Table 2: Data on Language Prevalence as per Ethnologue Website

Size of Native Speakers		Literacy Rate		Mobile Subscription Rate	
Urdu	68.6 million	Pakistan	54.9%	Pakistan	127 million
Hindi	260 m	India	62.8%	India	886 m
Tamil	67.8 m	Sri Lanka	91.2%	Sri Lanka	20 m
Bengali	189 m	Bangladesh	57.7%	Bangladesh	116.5 m
Punjabi	100 m	Pakistan	54.9%	Pakistan	127 m
Nepali	17 m	Nepal	57.4%	Nepal	21 m
Sinhala	16 m	Sri Lanka	91.2%	Sri Lanka	20 m

As on 2019 in the international context, there are 5.86 million articles in English, 5.36 million articles in Cebuano language, Svenska has got 3.74 million articles; Deutsch and Francis languages have got 2.31 and 2.11 million articles respectively. Languages from Dutch, Russian, Italian, Spanish, Polish, Waray, Vietnamese, Japanese, Chinese, and Portuguese are having more than one million articles.

As mentioned earlier, Asia has got 51% share of global Internet users. However, the numbers of Wikipedia articles in any of the South Asian languages are negligible in size in comparison to languages of developed countries. This indicates that telecom penetration and availability of mobile and computing devices may not sufficient for the normal users to participate in online field and contribute user-generated content. Beyond the infrastructural requirements, users' ability and skills in handling their respective mother tongues in the online arena is a challenging scenario that limiting the number of Wikipedia articles in major South Asian languages.

DATA FROM GOOGLE SEARCH TECHNIQUE

Popular search terms of 2015 and 2018 of Global News Trends were collected form Google Trends. To maintain commonality - search terms were converted into generic terms - Terrorism, Disaster, Health, Conflict, Politic, Strike, and World-cup etc. These generic terms were further translated into respective South Asian languages with the help of Google translate tool. Each translated terms were searched through Google. In each terms, language wise data sorted, language with highest search results was given 7 and lowest was given 1. Based on total of each language, the popularity is measured. Hindi language dominates with the highest score (32 in 2015 and 44 in 2019), followed by Nepali (26 & 36), and Urdu in third position (23 & 34). Other languages: Bengali (21 & 30), Tamil (15 & 26), Sinhala (13& 11) and Punjabi (10 & 15).

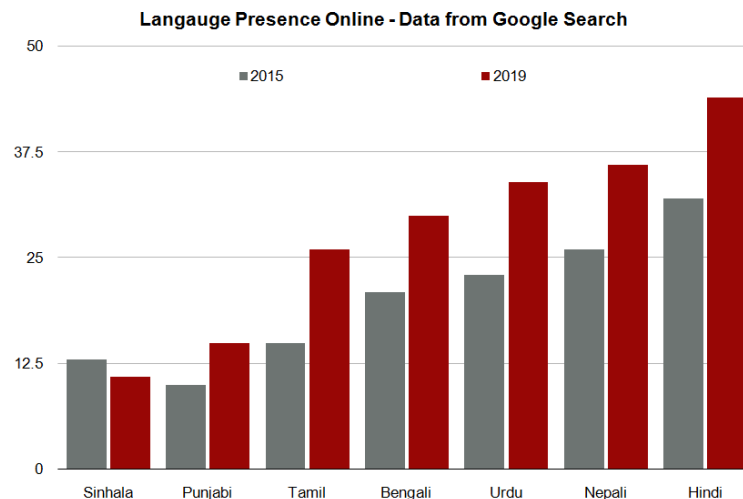


Chart 2: Source: Data Retrieved from Google Search

Google search results show Hindi language is top in the list, followed by Nepali and Urdu. That is in contrast to the Wikipedia data, where Urdu and Hindi dominate the list with more than one lakh articles followed by Tamil. Nepali is in second position in Google dataset whereas the same language is in low ranking in Wikipedia dataset, there could be the following reason for Nepali scoring high in Google - Hindi and Nepali are sharing the common scripts. Many results for Nepali terms, Google search results shown sites of Hindi news websites. Similarly, Bengali has scored more in Google than Wikipedia. Similar to Nepali language which shares common script with the Hindi language, Bengali script is commonly shared and used among many north-eastern languages of India such as Assamese and Bishnupriya.

Third Party on-line Language Measuring Formula

A internet based technology supported website - www.w3techs.com - provides services that includes the popularity of various languages which are available online. Though publicly this site didn't disclose how these languages are measured and ranked, however it indicated that like a search engine it uses a computer algorithm which crawls websites and retrieves data pertinent to languages used in the sampled websites. Data related to South Asian languages were retrieved from this technology based website, a ranking was created.

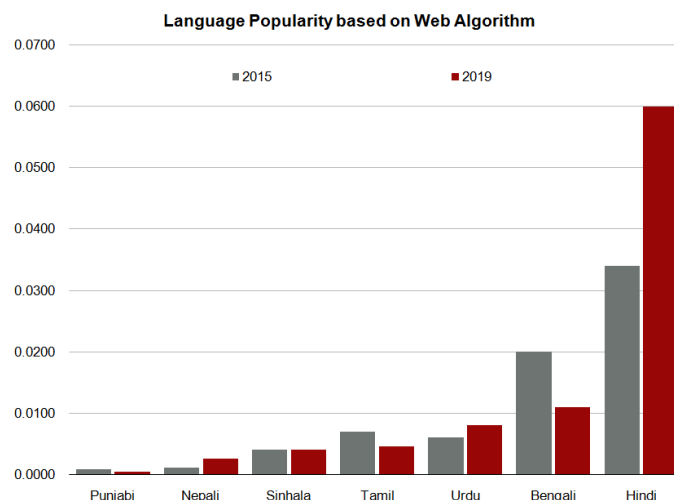


Chart 3: Source: Data Retrieved from www.w3techs.com

According to web algorithm dataset, Hindi language is on the top (0.034% in 2015 and 0.060 in 2019) followed by Bengali (0.02% & 0.011). Except Hindi and Bengali, all the remaining languages (in ranking order - Urdu, Tamil, Sinhala, Nepali and Punjabi) are having less than 0.001% of online content share. Globally, English language presence in online area is nearly 54% (it was 55% in 2015) and other popular languages such as Russian, German, and Spanish are having less than 6 % of web presence (it was less than 5% in 2015). Japanese language was in top five popular languages in this category in 2015, but in 2019 French replaced Japanese with 3.9% of web presence. Compare to English and other popular languages, South Asian languages are negligible in terms of its share in the online field. Positive development in this comparison between 2015 and 2019 dataset is that English domination is reducing and other languages are getting increasing presence. In South Asian context, only Hindi language has shown increasing trend.

Combined data on South Asian languages

Three different sets of data were retrieved with regard to South Asian languages, all these individual ranking were combined in order to take a holistic view on combined popularity of South Asian languages. In the final comparison, Hindi language has emerged as a leading languages which has higher presence in online, followed by the Urdu language. Bengali, Tamil, Nepali, Punjabi and Sinhala. The combined ranking of South Asian languages indicates that online field has got no boundary. Language, which is popular in one country, is equally popular in neighboring country, as noted earlier there is a commonality in sharing scripts between languages. Commonality and script sharing are two factors that influencing the online proliferation of South Asian languages.

Table 3: Combined Rank List

Combined Data of Rank List of South Asian Languages			
Wikipedia Rank	Google Rank	Algorithm Rank	Final Rank
Urdu	Hindi	Hindi	Hindi
Hindi	Nepali	Bengali	Urdu
Tamil	Urdu	Urdu	Bengali
Bengali	Bengali	Tamil	Tamil
Punjabi	Tamil	Sinhala	Nepali
Nepali	Punjabi	Nepali	Punjabi
Sinhala	Sinhala	Punjabi	Sinhala

DATA ANALYSIS AND INTERPRETATION

Out of eight countries in the South Asian region, the final list of languages which are popular in terms of three chosen indicators shows that these languages comes from five countries namely - India (all final list languages except Sinhala), Pakistan (Urdu and Punjabi), Bangladesh (Bengali), Nepal (Nepali) and Sri Lanka (Sinhala and Tamil). Languages from remaining three countries (Afghanistan, Bhutan, and Maldives) of South Asian region are not making it to the list. There could be many reasons for this condition: a) Languages of these three countries are not listed in any or one of these indicators - reasons could be that the number of respective speakers may be very marginal, hence the web algorithm doesn't included in it, and Google may not finding it viable in their services include these marginal languages and since there is no enough users in their respective languages that could be the reason for less or no articles in the Wikipedia. Size of some of the South Asian native speakers is much higher than any of languages of developed nation, which has got higher prevalence in the online field. Even after having sizeable native speakers, South Asian languages failed to make significant presence in online field. There may be multiple reasons for this scenario: 1] unlike developed

countries, computer software and mobile gadgets are not popularly customized into many South Asian languages. Eventually many mobile and computer users of South Asian region are forced to use these gadgets in English, which technically restricts active participation of native speakers. 2] Unicode based language fonts are Highly compatible with the digital gadgets and online space. Unicode based languages fonts are commonly available in many South Asian languages. Even after the presence of various Unicodebased languages, insignificant presence of South Asian languages indicates that users of computing and mobile gadgets are not familiar with the Unicode based language keyboard typing. Most probably they are using English alphabets for their language - transliteration. 3] based on Wikipedia statistics, page views traffic data (Table 4) were retrieved for South Asian languages. Significant size of page view traffic came from developed regions for many languages, particularly from Europe and United States. This trend indicates that good numbers of diaspora people of native speakers of South Asian languages are accessing their respective language content. Diaspora people might be having better technological support and online accessibility, however they may be not having much familiarity with the writing skills of their native language. That might have prevented them to contribute substantially to their language contents.

Table 4: Page view Traffic

Page View Traffic Data for Wikipedia of South Asian Languages		
Language	Page View Traffic Jan 2016	Page View Traffic Sep 2018
Hindi	81.4% traffic comes from India 9.3% from United States	93.3% traffic comes from India 5.3% from United States
Urdu	48.9% comes from Pakistan 13.7% from United States 5% from Europe 4.2% from India	35.9% comes from Pakistan 24.2% from United States 16.8% from China 9.9% from India
Bengali	49.3% from Bangladesh 16.3% from India 15.2% from United States	38.4% from Bangladesh 33.6% from United States 16.4% from India
Tamil	62% from India 10.5% from Sri Lanka 9.7% from United States	75.4% from India 7.5% from United States 7.0% from Srilanka
Nepali	57.1% from Nepal 25.8% from United States 4.8% from India 4.4 from Canada	53.6% from Nepal 13.9% from United States 10.4% from China 6.4% from India
Sinhala	71.9% from Sri Lanka 8.6% from United States 3.1% from South Korea	80.5% from Sri Lanka 5.7% from China 4.2% from Germany
Punjabi	38.6% from India 29.9% from United States 15.5% from France	49.1% from India 16.7% from China 13.8% from United States

SUGGESTIONS AND RECOMMENDATIONS

Internet access in South Asian countries has reached a significant point, equally social media presence and its influence is growing steadily in this region. It denotes that this region has got a critical mass of online users, equally it needs to be pointed out here that the majority of the population are still far from the digital world. On an average half of the population in South Asian countries are active users of internet, however this huge of chunk of users size doesn't translate into significance presence of South Asian languages in the online area. The growing telecom infrastructure and presence of young adults in this region are promising a productive future for online transformation of South Asian languages. Institutions and individuals collectively could make effort to create online content in regional languages. Some of the

possible potential opportunities in the long-term perspectives:

- Primary and secondary school teachers may be trained to teach usage of media and information tools in their respective learning environments. This would enable the teachers to inculcate the skills among the young school students to access better and reliable information resources. Access to genuine information through reliable media is a significant requirement in this information revolution era.
- Primary and secondary schools could be encouraged to establish media clubs, in which interested students may be encouraged to be part of it. Through these media clubs, students could be exposed to various kinds of media contents - normal news, motivated content and commercial messages. Exposure to these kinds of various content along with the intervention of teachers' inputs would help the students to develop required cognitive skills to interpret the media content in a rightful manner.
- National education boards could be encouraged to include Information and Communication Technology (ICT) based curriculum among the primary and secondary level students. That would enable the students to learn the basic digital media skills. In the senior secondary school level, a course on media literacy might be introduced that could help the senior students to deal with the onslaught of various media outlets.
- These media and information based interventions would enable the young population is having better understanding on accessing right information, understanding and interpreting media and information contents appropriately and it will enable young minds to participate in the various media platforms responsibly.

Notes:

- Table 1 data retrieved from <https://www.internetworldstats.com/stats3.htm#asia> in May 2019
- Table 2 data retrieved from <https://www.ethnologue.com> in May 2019
- Table 4 data retrieved from
- <https://stats.wikimedia.org/wikimedia/squids/SquidReportPageViewsPerLanguageBreakdown.htm> in May 2019
- Google trends data retrieved from <https://trends.google.com/trends/?geo=US> in March 2015 and May 2019
- Web algorithm data collected from this URL - http://w3techs.com/technologies/overview/content_language/all in March 2015 and May 2019

REFERENCES

1. D'Monte, L. (2014). 'Need more local language content for Internet to bloom in India.' *Mint newspaper*, *LiveMint.com*. Retrieved from <http://www.livemint.com/Opinion/zCFoFUXebEbxBOVq8a2UUJ/Need-more-local-language-content-for-Internet-to-bloom-in-In.html> (May, 2015).
2. Film Federation Of India. (2018). Retrieved from <http://filmfed.org/IFF2017.HTML>
3. Gerrand, P. (2007). *Estimating linguistic diversity on the Internet: A taxonomy to avoid pitfalls and paradoxes*. *Journal of Computer-Mediated Communication*, 12:1298–1321. doi: 10.1111/j.1083-6101.2007.00374.x
4. Indian Readership Survey. (2017). Retrieved from <http://mruc.net/uploads/posts/a27e6e912eedeab9ef944cc3315fba15.pdf> (May, 2019).

5. Mikami et al. (2005). *Language diversity on the Internet: An Asian view*. In Paolillo, J. et al., (Eds). *Measuring Linguistic Diversity on the Internet*, pp. 91-103. Retrieved from http://eprints.utm.my/3407/2/Measuring_Linguistic_Diversity.pdf (May, 2015).
6. Selvan, Arul. (2015). *Measuring Linguistic Diversity in Indian Online Scenario* in Jagtar, S. (2015). *Media and Information Literacy for Sustainable Development Goals*(1st ed., pp. 207-218). Goteberg: *The International Clearinghouse on Children, Youth and Media*, Nordicom.
7. Usman, J., Mustafa, L. J., & AGU, M. N. (2013). *Proverb as Determinant of Cultural Identity: The Imperative of the Three Regional Languages in Nigeria*. *IMPACT: International Journal of Research in Humanities, Arts and Literature*, 1, 47-58.
8. *South Asia most diverse with 650 languages*. (2018). Retrieved from <https://www.thehindu.com/news/national/karnataka/south-asia-most-diverse-with-650-languages/article22399276.ece>
9. *TRAI Annual Report*. (2018). Retrieved from https://main.trai.gov.in/sites/default/files/Annual_Report_21022019.pdf
10. *Why marketers should focus on Indian language content for consumers*. (2018). Retrieved from <https://economictimes.indiatimes.com/small-biz/marketing-branding/marketing/why-marketers-should-focus-on-indian-language-content-for-consumers/articleshow/66049831.cms?from=mdr>

